UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/622,526	07/18/2003	Yufuko Hidaka	JP920020117US1	2803

25259 7590 04/10/2007	EXAMINER	
IBM CORPORATION	MYINT, DENNIS Y	
3039 CORNWALLIS RD.		
DEPT. T81 / B503, PO BOX 12195	ART UNIT	PAPER NUMBER
REASEARCH TRIANGLE PARK, NC 27709	2162	

SHORTENED STATUTORY PERIOD OF RESPONSE	NOTIFICATION DATE	DELIVERY MODE
3 MONTHS	04/10/2007	ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

If NO period for reply is specified above, the maximum statutory period will apply and will expire 6 MONTHS from the mailing date of this communication.

Notice of this Office communication was sent electronically on the above-indicated "Notification Date" and has a shortened statutory period for reply of 3 MONTHS from 04/10/2007.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

RSWIPLAW@us.ibm.com

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE _3_ MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on _07/18/2007_.

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) _1-22_ is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _1-22_ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on _07/18/2007_ is/are: a)☒ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☒ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☒ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date _05/25/2006_.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____ .

5)☐ Notice of Informal Patent Application

6)☐ Other: _____.

DETAILED ACTION

1.   Claims 1-22 have been examined.

Claim Rejections - 35 USC § 112

2.   The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly
> claiming the subject matter which the applicant regards as his invention.

3.   Claim 11-13 and 14-15 are rejected under 35 U.S.C. 112, second paragraph, as

being indefinite for failing to particularly point out and distinctly claim the subject matter

which applicant regards as the invention.

As per claim 11, the claim in lines 9-10 recites *"in a manner useful to a user"*.

Said limitation is vague, thereby rendering the scope of the claim unascertainable. As

such claim 11 is rejected under 35 U.S.C. 112 Second Paragraph.

Claims 12 and 13 are rejected under 35 U.S.C. 112 Second Paragraph because

said claims depend on claim 11.

As per claim 14, the claim in lines 10 recites *"useful information"*. Said limitation

is vague, thereby rendering the scope of the claim unascertainable. As such claim 11 is

rejected under 35 U.S.C. 112 Second Paragraph.

Claim s 15 is rejected under 35 U.S.C. 112 Second Paragraph because said

claims depend on claim 14.

Claim Rejections - 35 USC § 101

4.   35 U.S.C. 101 reads as follows:

Whoever invents or discovers any new and useful process, machine, manufacture, or composition of matter, or any new and useful improvement thereof, may obtain a patent therefor, subject to the conditions and requirements of this title.

5.   Claims 1-15 and 18-22 are rejected under 35 U.S.C. 101 because the claimed invention is directed to non-statutory subject matter.

Claims 1-6 are directed to an information collection system. According Paragraph 0035 of the specification, said information collection is interpreted as a system of software functions. Paragraph 0035 of the specification recites *It is possible to take only the web applications server 20 as an information collection system in a narrow sense. As used herein, the work "system" includes functions in the same housing or connected via a prescribed network"*. Therefore, said information collection system is software per se and does not fall in the four statutory categories.

Claims 7-10 are directed to an application server. According to the specification, said *application server* is software per se. Paragraph 0016 of the specification recites *"An application server in accordance with the present invention comprises a section for receiving information about a user's interest; a section for acquiring HTML documents from a plurality of sites based on the information received from the request receiving section; a vocabulary information processing mechanism for reading an ontology based on the information received from the user request receiving section to acquire vocabulary information; and an information identifying section for extracting data objects with respect to the HTML documents acquired from the HTML acquiring section, relying*

on tags of the HTML documents and based on the vocabulary information offered from the vocabulary information processing mechanism". As such, said application server is software per se and does not fall with the four statutory categories.

Claim 11 in lines 7 recites *"extracting **useful** information"*. Said limitation is relative in nature, that is, extracted information might be useful to one person and might not be useful to another. Results of said step is therefore not concrete. Said step of extracting useful information cannot be substantially repeatable. Therefore, claim 11 is not statutory and is rejected under 35 U.S.C. 101. Claims 12 and 13 depend on claim 11 and are therefore rejected under 35 U.S.C. 101.

In addition, claim 11 fails to produce tangible results and therefore is not statutory. This claimed subject matter lacks a practical application of a judicial exception (law of nature, abstract idea, naturally occurring article/phenomenon) since it fails to produce a tangible result. Specifically, the claimed subject matter does not produce a tangible result because the claimed subject matter fails to produce a result that is limited to having real world value rather than a result that may be interpreted to be abstract in nature as, for example, a thought, a computation, or manipulated data. More specifically, the claimed subject matter provides for "extracting necessary information from said plurality of data files based on results from said means for analyzing". This produced result remains in the abstract and, thus, fails to achieve the required status of having real world value.

Claims 12 and 13 depend on claim 11 and are therefore rejected under 35 U.S.C. 101.

As per claim 14 and 15, the claims are directed to an information collection method. However, said claims fails to produce tangible results. This claimed subject matter lacks a practical application of a judicial exception (law of nature, abstract idea, naturally occurring article/phenomenon) since it fails to produce a tangible result. Specifically, the claimed subject matter does not produce a tangible result because the claimed subject matter fails to produce a result that is limited to having real world value rather than a result that may be interpreted to be abstract in nature as, for example, a thought, a computation, or manipulated data. More specifically, the claimed subject matter provides for "extracting necessary information from said plurality of data files based on results from said means for analyzing". This produced result remains in the abstract and, thus, fails to achieve the required status of having real world value.

Additionally, Claim 14 in lines 10 recites *"extracting **useful** information"*. Said limitation is relative in nature, that is, extracted information might be useful to one person and might not be useful to another. Results of said step is therefore not **concrete**. Said step of extracting useful information cannot be substantially repeatable. Therefore, claim 14 is not statutory and is rejected under 35 U.S.C. 101. Claims 15 depends on claim 14 and are therefore rejected under 35 U.S.C. 101.

Claims 18-22 is directed to a program product and therefore the claimed subject matter does not fall in the four statutory categories. This claimed subject matter lacks a practical application of a judicial exception (law of nature, abstract idea, naturally occurring article/phenomenon) since it fails to produce a tangible result. Specifically, the claimed subject matter does not produce a tangible result because the claimed subject matter fails to produce a result that is limited to having real world value rather than a

result that may be interpreted to be abstract in nature as, for example, a thought, a

computation, or manipulated data. More specifically, the claimed subject matter

provides for "summing up the analyzed information". This produced result remains in the

abstract and, thus, fails to achieve the required status of having real world value.

Claim Rejections - 35 USC § 103

6.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

7.   This application currently names joint inventors. In considering patentability of

the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of

the various claims was commonly owned at the time any inventions covered therein

were made absent any evidence to the contrary. Applicant is advised of the obligation

under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was

not commonly owned at the time a later invention was made in order for the examiner to

consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g)

prior art under 35 U.S.C. 103(a).

8.   Claim 1-4, 6, 7, 21, and 22 rejected under 35 U.S.C. 103(a) as being

unpatentable over Craven et al., (hereinafter "Craven") Learning to Extract Symbolic

Knowledge form the World Wide Web (Proceedings of the 15the National Conference

on Artificial Intelligence (AAAI), 1998, in view of Shanahan et al., (U.S. Patent

Application Publication Number 2005/0022114).

As per claim 1, Craven is directed to an information collection system (Craven, *Abstract: The goal of the research described here is automatically create a computer understandable world wide web knowledge base whose content mirrors that of the World Wide Web*) and teaches the limitations:

"means for acquiring a plurality of data files via a network" (Craven, Page 8, Column 2, Lines 30-37, i.e., *In this section, we describe an experiment that involves evaluating our approach in a novel, on-line environment. We have developed a Web-Crawling system that populates a knowledge base with class and relation instances as it explores the Web*);

""means for analyzing the plurality of data files acquired using an ontology of relational description of terms" (Craven, Page 1, Column 2, Lines 6-16, i.e., *The approach explored in our research is to develop a trainable system that can be taught to extract various types of information by automatically browsing the web. This system accepts two inputs: 1)* **an ontology** *specifying the classes and relations of interest. An example of such an ontology is provided in the top half of Figure 1.;* and Craven, Figure 1.);

"and means for extracting necessary information from said plurality of data files based on results from said means for analyzing" (Craven, Page 1, Column 2, Lines 6-16, i.e., *The approach explored in our research is to develop a trainable system that can be taught* **to extract** *various types of information by automatically browsing the web.;* Page 6 Column 2 Lines 37 through Page 7 Column 2 Line 26, i.e., *Extract Text Fields*).

Craven does not explicitly teach the limitation: "using a prescribed extraction rule"

On the other hand, Shanahan teaches the limitation:

"using a prescribed extraction rule" (Shanahan, Paragraph 0230, i.e., *The entity*

*extractor 1802 **includes generic rules** for extracting entity types such as names of*

cities, people, products, dates, noun phrases, etc. These generic rules do not specify

entities per se. Instead they specify generic entities that are capable of detecting that a

capitalized noun is likely to be a person's name rather than a name of a city, which can

be performed by using the context surrounding the identified noun). Note that the

system of Shanahan first analyzes before extracting using generic rules. Also note that

the system of Shanahan teaches employing ontology to categorize extracted results in

Paragraph 0428.

At the time the invention was made, it would have been obvious to a person of

ordinary skill in the art to add the feature of using a prescribed extraction rule as taught

by Shanahan to the system of Craven so that the resultant system would also comprise

using a prescribed extraction rule when analyzing acquired files/documents. One would

have been motivated to do so in order to exclude irrelevant results from a corpus of

files/document. Using extraction rules to extract document/files is well known in the art.

As per claim 2, Craven in view of Shanahan teaches the limitations:

"wherein said means for acquiring obtains documents described in Hypertext

Markup Language using Uniform Resource Locators based on a user's interest"

(Craven, Page 8, Column 2, Lines 30-37, i.e., *In this section, we describe an experiment*

that involves evaluating our approach in a novel, on-line environment. We have

developed a Web-Crawling system that populates a knowledge base with class and

relation instances as it explores the Web; Note that Hyper Text Markup Language and

URL are already inherent said teaching of craven), and "said means for analyzing

analyzes said documents using specific tag information." (Craven, Page 8, Column 2,

Lines 30-37, i.e., when web crawlers explore the Web, HTML tags are always analyzed;

Also, See Paragraph 0307 of Shanahan which teaches as *That is, every page that is*

viewed on the browser during a current session (e.g., starting from a first identified

document) is analyzed and all strings that are marked up (e.g., everything between the

HTML <a> and) are stored by a plug-in with an expanded URL (e.g., base URL

plus relative URL found in the `href` field within the <a> tag)).

As per claim 3, the claim recites in Line 3-4 as *catalog and/or specification*

information that are put into a values. Examiner chooses said limitation to mean *catalog*

or specification information that are put into a values. As such, Shanahan teaches the

limitations of claim 3 as follows:

"wherein said prescribed extraction rule used in the analysis by said means for

analyzing is features constituting catalog or specification information that are put into a

rule" (Shanahan, Paragraph 0230, i.e., *The entity extractor 1802 includes generic rules*

for extracting entity types such as names of cities, people, products, dates, noun

phrases, etc. These generic rules do not specify entities per se. Instead they specify

generic entities that are capable of detecting that a capitalized noun is likely to be a

person's name rather than a name of a city, which can be performed by using the

context surrounding the identified noun).

As per claim 4, Craven teaches the limitation:

"wherein said means for analyzing analyzes contents transversely using the

ontology with respect to the plurality of data files using different terms" (Craven, Page 1,

Column 2, Lines 6-16, i.e., *The approach explored in our research is to develop a*

trainable system that can be taught to extract various types of information by

automatically browsing the web. This system accepts two inputs: 1) **an ontology**

specifying the classes and relations of interest. An example of such an ontology is

provided in the top half of Figure 1.; and Craven, Figure 1.).

As per claim 6, Craven teaches the limitations:

"further comprising ontology storing means for storing ontologies that differ per

object, wherein said means for analyzing conducts an analysis by reading a prescribed

ontology from said ontology storing means" (Craven, Page 1, Column 2, Lines 6-16, i.e.,

The approach explored in our research is to develop a trainable system that can be

taught to extract various types of information by automatically browsing the web. This

system accepts two inputs: 1) **an ontology** *specifying* **the classes** *and relations of*

interest. An example of such an ontology is provided in the top half of Figure 1.; and

Craven, Figure 1.).

Claim 7 is rejected on the same basis as claim 2. Claim 2 incorporates all the limitations of claim 1 and claim 2.

Claim 21 is rejected on the same basis as claim 1.

Claims 22 is rejected on the same basis as claim 7. Claim 7 was rejected on the same basis of claim 2 and Claim 2 incorporates all the limitations of claim 1 and claim 2.

9.   Claim 5, 8, 10-12, 14-18, and 20 rejected under 35 U.S.C. 103(a) as being unpatentable over Craven in view of Shanahan and further in view of Julien (U.S. Patent Number 6694307).

As per claim 5 Craven in view of Shanahan does not explicitly teach the limitation: "further comprising offering means for reconstructing the information extracted by said extracting means, aggregating equivalent relationships from said information, and offering them to a user terminal."

On the other hand, Julien teaches the limitation:

"further comprising offering means for reconstructing the information extracted by said extracting means, aggregating equivalent relationships from said information and offering them to a user terminal" (Julien, Column 4 Lines 22-25, i.e., *The aggregator unit relies on pre-determined clustering rules to correlate and establish relationships between the information elements identified in each web page;* and Column 10 Lines 39-47, i.e., *5. The aggregator unit 34 processes the extracted information elements in order to correlate and establish relationships between the information elements*

*identified in each Web page, thus compiling an aggregated list of contact information for each Web page. The aggregator unit next combines **the set of aggregated lists** into a final, **structured list** containing the requested contact information, removing any similar contact information and combining any complementary contact information. 6. The aggregator unit 34 transmits the final list containing the requested contact information to the system user, **displayed to the system user in a data display screen** on the monitor of the workstation 12).*

At the time the invention was made, it would have been obvious to a person of ordinary skill in the art to add the feature of aggregating equivalent relationships in document/file extraction, as taught by Julien, to the system of Craven in view of Shanahan so that the resultant system would also comprise aggregating equivalent relationships in document/file extraction. One would have been motivated to do so in order to *provide a novel system for extracting and structurally compiling specific information from unstructured digitized data, such as the Web pages of the WWW"* (Julien, Column 2 Lines 12-15).

As per claim 8, Julien teaches the limitation:

"further comprising an extraction rule processing mechanism that offers an extraction rule for applying extraction processing to said HTML documents acquired from said HTML acquiring section, wherein said extracting position information identifying section obtains the extraction data objects based on the extraction rule offered from said extraction rule processing mechanism" (Julien, Column 7 Lines 35-53,

i.e., *In this particular example of implementation, the identification unit 30 includes a tagging unit 36 operative to mark the data of a particular Web page by inserting tags, in order to identify information elements. Each tag, in addition to identifying a particular information element, includes data indicative of a category for the identified information element, as will be described below. More specifically, when the identification unit 30 recognises in the data of a particular Web page an information element that is relevant to contact information, the tagging unit 36 inserts into the data of the Web page a machine-readable tag, such as digitized text or symbol(s). The tag is inserted at a predetermined **position** with respect to the identified information element, in one example immediately adjacent and preceding the information element. An example of such a tagging unit is the InfoTagger (trade-mark), manufactured by the Canadian National Research Council (CNRC) located in Ottawa, Canada).*

As per claim 10, Shanahan in view of Julien teaches the limitations:

"an information arranging/aggregating section for applying a summing process to the plurality of extraction data objects obtained by said extracting position information identifying section" (Shanahan, Paragraph 0260, *i.e., or alternatively to the entire document, **from which the summary was derived**. The similarity measure can be performed using either resulting form. In this equation a similarity measure is generated for each entity (represented as the entity plus a context) and result document "doc" (**represented as a summary** or the entire document content).);*

"a summing result object producing section for producing a table and/or a list

based on a result of the summing process by said information arranging/aggregating

section" (Shanahan, Paragraph 0260, i.e., i.e., *Also in the equation, "doc" refers to*

either the document summary that appears (as an element in a result list) *in the*

results page of the service or alternatively to the entire document, from which the

summary was derived. The similarity measure can be performed using either resulting

form. In this equation a similarity measure is generated for each entity (represented as

the entity plus a context) and result document "doc" (represented as a summary or the

entire document content));

"and a user request transmitting section for transmitting a summing result object

produced by said summing result object producing section" (Julien, Column 4 Lines 22-

25, i.e., *The aggregator unit 34 transmits the final list containing the requested contact*

information to the system user, ***displayed to the system user in a data display***

screen *on the monitor of the workstation 12*).

Claim 11 is rejected on the same basis as claim 5. Claim 5 incorporates all the

limitations of claim 11 and claim 5.

As per claim 12, Craven in view of Shanahan teaches the limitations:

"further comprising a step of extracting information from HTML documents

acquired via the network based on table tags or list tags" (Craven, Page 8, Column 2,

Lines 30-37, i.e., when web crawlers explore the Web, HTML tags are always analyzed;

Also, See Paragraph 0307 of Shanahan which teaches as *That is, every page that is viewed on the browser during a current session (e.g., starting from a first identified document) is analyzed and all strings that are marked up (e.g., everything between the HTML <a> and) are stored by a plug-in with an expanded URL (e.g., base URL plus relative URL found in the `href` field within the <a> tag)).*

Claim 14 is rejected on the same basis claim 11.

As per claim 15, Craven in view of Shanahan and further in view of Julien teaches the limitation:

"wherein said information about the user's interest is information relating to object URLs and an ontology type, and the step of acquiring said plurality of documents acquires HTML documents based on said object URLs and extracts table portions or list portions from said HTML documents" (Craven, Page 1, Column 2, Lines 6-16, i.e., *The approach explored in our research is to develop a trainable system that can be taught to extract various types of information by automatically browsing the web. This system accepts two inputs: 1) **an ontology** specifying the classes and relations of interest. An example of such an ontology is provided in the top half of Figure 1.;* and Craven, Figure 1; Craven, Page 8, Column 2, Lines 30-37, i.e., *In this section, we describe an experiment that involves evaluating our approach in a novel, on-line environment. We have developed a Web-Crawling system that populates a knowledge base with class and relation instances as it explores the Web;* Note that Hyper Text Markup Language and URL are already inherent said teaching of craven; and Craven, Page 8, Column 2,

Lines 30-37, i.e., when web crawlers explore the Web, HTML tags are always analyzed;

Also, See Paragraph 0307 of Shanahan which teaches as *That is, every page that is*

viewed on the browser during a current session (e.g., starting from a first identified

document) is analyzed and all strings that are marked up (e.g., everything between the

HTML <a> and) are stored by a plug-in with an expanded URL (e.g., base URL

plus relative URL found in the `href` field within the <a> tag)).

As per claim 16, Craven in view of Shanahan, further in view of Julien teaches

the limitations:

"an information collection method comprising: acquiring a plurality of Web pages

including information expressed by different vocabularies with respect to associated

contents, respectively" (Craven, Page 8, Column 2, Lines 30-37, i.e., *In this section, we*

describe an experiment that involves evaluating our approach in a novel, on-line

environment. We have developed a Web-Crawling system that populates a knowledge

base with class and relation instances as it explores the Web);

"extracting information from said plurality of acquired Web pages based on table

tags or list tags" (Craven, Page 1, Column 2, Lines 6-16, i.e., *The approach explored in*

*our research is to develop a trainable system that can be taught **to extract** various*

types of information by automatically browsing the web.; Page 6 Column 2 Lines 37

through Page 7 Column 2 Line 26, i.e., *Extract Text Fields;* and Craven, Page 8,

Column 2, Lines 30-37, i.e., when web crawlers explore the Web, HTML tags are

always analyzed; Also, See Paragraph 0307 of Shanahan which teaches as *That is,*

every page that is viewed on the browser during a current session (e.g., starting from a first identified document) is analyzed and all strings that are marked up (e.g., everything between the HTML <a> and) are stored by a plug-in with an expanded URL (e.g., base URL plus relative URL found in the `href` field within the <a> tag));

"analyzing said extracted information transversely with respect to the different vocabularies of said plurality of Web pages based on an ontology representing relationships between vocabularies" (Craven, Page 1, Column 2, Lines 6-16, i.e., *The approach explored in our research is to develop a trainable system that can be taught to extract various types of information by automatically browsing the web. This system accepts two inputs: 1)* **an ontology** *specifying the classes and relations of interest. An example of such an ontology is provided in the top half of Figure 1.;* and Craven, Figure 1.);

."summing up the analyzed information; and transmitting a summing result to a user terminal" (Shanahan, Paragraph 0260, i.e., i.e., *Also in the equation, "doc" refers to* **either the document summary that appears (as an element in a result list)** *in the results page of the service or alternatively to the entire document,* **from which the summary was derived**. *The similarity measure can be performed using either resulting form. In this equation a similarity measure is generated for each entity (represented as the entity plus a context) and result document "doc" (represented as a summary or the entire document content)).*

As per claim 17, the claim in Line 2-4 recites "superordinate/subordinate

conceptual processing and/or relational processing". Examiner chooses this limitation to

mean "relational processing". As such, Craven in view Shanahan teaches the limitation:

"wherein said summing" (Shanahan, Paragraph 0260, i.e., i.e., *Also in the*

equation, "doc" refers to **either the document summary that appears (as an element**

in a result list) *in the results page of the service or alternatively to the entire document,*

from which the summary was derived. *The similarity measure can be performed*

using either resulting form. In this equation a similarity measure is generated for each

entity (represented as the entity plus a context) and result document "doc" (represented

as a summary or the entire document content)) "applies relational processing to the

different vocabularies on the respective Web pages to implement matching of items"

(Craven Figure 1, i.e., *Ontology*).

Claim 18 is rejected on the same basis as claim 11.

As per claim 20, Craven in view of Shanahan, further in view of Julien teaches

the limitations:

"wherein said function of reconstructing processes an equivalent relationship with

respect to associated vocabulary" (Julien, Column 4 Lines 22-25, i.e., *The aggregator*

*unit relies on pre-determined clustering rules to correlate and establish **relationships***

between the information elements identified in each web page; and Column 10 Lines

39-47, i.e., *5. **The aggregator unit** 34 processes the extracted information elements in*

order to correlate and establish relationships between the information elements

identified in each Web page, thus compiling an aggregated list of contact information for

*each Web page. The aggregator unit next combines **the set of aggregated lists** into a*

*final, **structured list** containing the requested contact information, removing any similar*

contact information and combining any complementary contact information. 6. The

aggregator unit 34 transmits the final list containing the requested contact information to

*the system user, **displayed to the system user in a data display screen** on the*

monitor of the workstation 12; and Craven, Page 5 Column 2 Lines 27 through Page 6

Column 2 Line 34, i.e., ***Reorganizing Relations Instances***) and "value to insert a new

relationship to reconstruct the information" (Craven, Page 6 Column 2 Lines 21-26, i.e.,

In this case, the algorithm has learned to exploit the face that departments often have a

page that serves as a graduate student directory, and that any student whose home

page is pointed to by this directory is a member of the department. Also note Figure 1 of

Craven).

10.   Claim 9   rejected under 35 U.S.C. 103(a) as being unpatentable over Craven in

view of Shanahan and further in view of Hayashi (U.S. Patent Number 5628008).

As per claim 9, Craven in view of Shanahan does not explicitly teach the

limitation: "further comprising an inference processing mechanism for executing an

inference operation based on an axiom rule, wherein said extracting position information

identifying section obtains the extraction data objects based on the inference operation

executed by said inference processing mechanism".

On the other hand, Hayashi teaches the limitation:

"further comprising an inference processing mechanism for executing an inference operation based on an axiom rule" (Hayashi, Figure 10, i.e., *INFERENCE SEC. 22;* and Column 10 Lines 40-46, i.e., *the inference section 22;*), "wherein said extracting position information identifying section obtains the extraction data objects based on the inference operation executed by said inference processing mechanism" (Column 12 Line 66 through Column 13 29, i.e., *FIG. 13 is a diagram illustrating an example of a substructure of a structured document for explaining a specific example of correcting a search formula, and FIG. 14 is a diagram illustrating an example of the structure of a search formula subject to correction. A case is considered in which retrieval is effected by a search formula having the structure including a portion indicated by the broken lines in FIG. 14 with respect to the structured document having the structure shown in FIG. 13. **The * marks in the drawing indicate extracting positions.** In this case, the search result is the node a "Although it may appear that the introduction was not interesting, . . . " in FIG. 13. Here, the user points out that the node b "First, . . . " in FIG. 13 should also appear as the search result. Then, the structure analyzing section 23 makes a comparison between the structure of the search formula shown in FIG. 14 and a surrounding substructure including the node b in FIG. 13, and extracts a branch which is not present in the vicinity of the node b and is present only in the structure of the search formula. In this example, since a branch k indicated by the broken lines in FIG. 14 is not present in the vicinity of the node b in FIG. 13, this branch k is extracted by the structure analyzing section 23. Then, **if the inference section 22***

allows the branch k to be deleted from the structure of the search formula, the search-

formula correction processing section 21 deletes the branch k from the structure of the

search formula, so that the search formula is corrected in such a manner as to be

provided with only the structure indicated by the solid lines in FIG. 14. Then, as a

search is carried out by the corrected search formula, both portions of the node a and

the node b in FIG. 13 are included in the search result). This disclosure of Hayashi

teaches that inference section 22 decides extraction positions.

At the time the invention was made, it would have been obvious to a person of

ordinary skill in the art to add the feature of having inference section/engine determine

extracting positions, as taught by Hayashi, to the application server of Craven in view of

Shanahan so the resultant application server would also comprise he feature of having

inference section/engine determine extracting positions. One would have been

motivated to do so in order to provide a search mechanism for effecting retrieval

processing with respect to a structured document (Hayashi, Column 1 Lines 52-55).

11.   Claim 19   rejected under 35 U.S.C. 103(a) as being unpatentable over Craven in

view of Shanahan, further in view of Julien and further in view of Hayashi (U.S. Patent

Number 5628008).

As per claim 19, Craven in view of Shanahan and further in view of Julien does

not explicitly teach the limitation: "wherein said function of analyzing executes an

inference operation based on a prescribed axiom rule".

On the other hand, Hayashi teaches the limitation:

"wherein said function of analyzing executes an inference operation based on a prescribed axiom rule" (Hayashi, Figure 10, i.e., *INFERENCE SEC. 22;* and Column 10 Lines 40-46, i.e., *the inference section 22;*).

At the time the invention was made, it would have been obvious to a person of ordinary skill in the art to add the feature of having inference section/engine determine extracting positions, as taught by Hayashi, to the application server of Craven in view of Shanahan and further in view of Julien so the resultant application server would also comprise he feature of having inference section/engine determine extracting positions. One would have been motivated to do so in order to provide a search mechanism for effecting retrieval processing with respect to a structured document (Hayashi, Column 1 Lines 52-55).

Allowable Subject Matter

12.   Claims 13 is objected to as being dependent upon a rejected base claims, but would be allowable if rewritten in independent form including all of the limitations of the base claims and any intervening claims, assuming correction of the claim rejections under 35 U.S.C. 101 above.

The following is a statement of reasons for the indication of allowable subject matter. Referring to claims 13, Craven in view of Shanahan and further in view of either Julien or Hayashi fails to teach what claim 13 of the claimed invention recites that "wherein said step of analyzing comprises a step of performing positioning of a table using said extraction rule that is obtained by putting features constituting catalog and/or

specification information into a rule, and a step of smoothing a swing of vocabularies

based on said ontology that defines vocabulary information representing whether or not

a headline of the positioned table is a vocabulary that is generally used in a category

designated by the user".

Therefore, claim 13 is allowable if written in an independent form.

Conclusion

13.   The prior art made of record and not relied upon is considered pertinent to

applicant's disclosure as follows.

U.S. Patent Application Publication Number 2003/0115189 (Srinivasa et al.,)

14.   Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Dennis Myint whose telephone number is (571) 272-

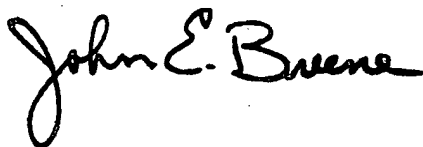5629. The examiner can normally be reached on 8:30AM-5:30PM Monday-Friday.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, John Breene can be reached on (571) 272-4107. The fax phone number for

the organization where this application or proceeding is assigned is 571-273-5629.

Information regarding the status of an application may be obtained from the

Patent Application Information Retrieval (PAIR) system. Status information for

published applications may be obtained from either Private PAIR or Public PAIR.

Status information for unpublished applications is available through Private PAIR only.

For more information about the PAIR system, see http://pair-direct.uspto.gov. Should

you have questions on access to the Private PAIR system, contact the Electronic

Business Center (EBC) at 866-217-9197 (toll-free).

/Hpham/   Dennis Myint
                                                                           Examiner
                                                                           AU-2162

JOHN BREENE
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100